

AN INEQUALITY FOR RELATIVE ENTROPY AND LOGARITHMIC SOBOLEV INEQUALITIES IN EUCLIDEAN SPACES

KATALIN MARTON

Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences

June 18, 2012

ABSTRACT. For a class of density functions $q(x)$ on \mathbb{R}^n we prove an inequality between relative entropy and the weighted sum of conditional relative entropies of the following form:

$$D(p||q) \leq \text{Const.} \sum_{i=1}^n \rho_i \cdot D(p_i(\cdot|Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n) || Q_i(\cdot|Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n))$$

for any density function $p(x)$ on \mathbb{R}^n , where $p_i(\cdot|y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n)$ and $Q_i(\cdot|x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ denote the local specifications of p resp. q , and ρ_i is the logarithmic Sobolev constant of $Q_i(\cdot|x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$. Thereby we derive a logarithmic Sobolev inequality for a weighted Gibbs sampler governed by the local specifications of q . Moreover, the above inequality implies a classical logarithmic Sobolev inequality for q , as defined for Gaussian distribution by L. Gross. This strengthens a result by F. Otto and M. Reznikoff. The proof is based on ideas developed by F. Otto and C. Villani in their paper on the connection between Talagrand's transportation-cost inequality and logarithmic Sobolev inequality.

¹This work was supported in part by the grants OTKA T-32323 and K-76088 of the Hungarian Academy of Sciences

²AMS 2000 subject classifications: 52A40, 60K35, 82B20, 82C22.

³Key words and phrases: Relative entropy, Wasserstein distance, Fokker-Planck equation, gradient flow, non-compact spin system, Gibbs sampler, weakly dependent random variables, logarithmic Sobolev inequality, transportation-cost inequality.

⁴Subject classification: 82C22, 60J05, 35Q84, 60J25, 82B21.

1. Introduction.

The motivation for this paper was to prove logarithmic Sobolev inequalities on product spaces, under possibly general conditions.

First we define some basic concepts:

Definition. For probability measures p and q on \mathbb{R}^m ($m \geq 1$ integer), we denote by $D(p||q)$ the relative entropy of p with respect to q :

$$D(p||q) = \int_{\mathbb{R}^m} \log \frac{dp(u)}{dq(u)} dp(u) \quad \text{if } p \ll q, \quad (1.1)$$

and ∞ otherwise. We always have in mind probability measures absolutely continuous with respect to the Lebesgue measure, and denote by the same letter their density functions. If p and q are density functions on \mathbb{R}^m then

$$D(p||q) = \int_{\mathbb{R}^m} p(u) \log \frac{p(u)}{q(u)} du \quad \text{if } p \ll q, \quad (1.2)$$

and ∞ otherwise. If Z and U are random variables with values in \mathbb{R}^m and distributed according to $p = \mathcal{L}(Z)$ resp. $q = \mathcal{L}(U)$, then we shall also use the notation $D(Z||U)$ for the relative entropy $D(p||q)$.

Definition. For measures p and q on \mathbb{R}^m , the Fisher information of p with respect to q is defined as

$$I(p||q) = \int_{\mathbb{R}^m} \left| \nabla \log \frac{p(u)}{q(u)} \right|^2 p(du), \quad (1.3)$$

if $\log(p(u)/q(u))$ is smooth.

Definition. The distribution q on \mathbb{R}^m satisfies a logarithmic Sobolev inequality with constant ρ if

$$D(p||q) \leq \frac{1}{2\rho} \cdot I(p||q)$$

for all density functions p on \mathbb{R}^m with $\log(p(u)/q(u))$ smooth.

A logarithmic Sobolev inequality for a probability measure q is equivalent to the hypercontractivity of the diffusion semigroup associated with q . The prototype is Gross' logarithmic Sobolev inequality for Gaussian measure which is associated to the Ornstein-Uhlenbeck semigroup [1], [2]. Another use of logarithmic Sobolev inequalities is to derive transportation cost inequalities (a tool to prove measure concentration), c.f. F. Otto, C. Villani [3]. The logarithmic Sobolev inequality for the stationary distribution of a spin system is equivalent to the property called "exponential decay of correlation"; for this concept we refer to Bodineau and Helffer [4] and Helffer [5].

In Euclidean spaces of dimension greater than 1, no simple characterization is available for the measures q satisfying a logarithmic Sobolev inequality with some positive constant. A well-known sufficient condition was given by Bakry and Emery [6]: A density function $q(x) = \exp(-V(x))$ on \mathbb{R}^m satisfies a logarithmic Sobolev inequality provided V is uniformly strictly convex. Another useful result is Holley and Stroock's perturbation lemma [7] which asserts that if q and \tilde{q} are density functions on \mathbb{R}^m , such that the ratio $\tilde{q}(x)/q(x)$ is bounded both from above and below, then q and \tilde{q} either both satisfy a logarithmic Sobolev inequality, or neither of them does.

For measures on Euclidean spaces with non-compact support, it has been a challenging task to derive logarithmic Sobolev inequalities from logarithmic Sobolev inequalities for the local specifications. (The local specifications of the measure $q = \mathcal{L}(X_1, \dots, X_m)$ on \mathbb{R}^m are the conditional densities $Q_i(\cdot | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_m) = \mathcal{L}(X_i | X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}, \dots, X_m = x_m)$.) Let q be a density function on a Euclidean space, and assume that the local specifications of q satisfy logarithmic Sobolev inequalities with constants ρ_i . It has been clear for a long time that a reasonable approach to prove a logarithmic Sobolev inequality for q is to assume that the mixed partial derivatives of $V(x) = -\log q(x)$ are not too large relative to the numbers ρ_i . This line was followed by B. Zegarlinski [8] and, following in his footsteps, G. Royer [9], Théorème 5.2.1). Their results were improved by F. Otto and M. Reznikoff [10]. The present paper follows this line, too. The conditions of Otto and Reznikoff's main theorem helped to find the proper conditions for the results in the present paper, however, our approach is entirely different from theirs. We shall discuss Otto and Reznikoff's theorem at the end of Section 2.

2. Statement of the results

Let \mathbb{R}^N denote the N -dimensional Euclidean space equipped with the Euclidean distance and the Borel σ -algebra.

Let us fix a density function

$$q(x) = \exp(-V(x)), \quad x \in \mathbb{R}^N.$$

We shall use the following

Notation:

- q : a fixed density function on \mathbb{R}^N ;
- $X = (X_1, X_2, \dots, X_N)$: random sequence in \mathbb{R}^N , $\mathcal{L}(X) = q$;
- p : another density function on \mathbb{R}^N ;
- $Y = (Y_1, Y_2, \dots, Y_N)$: random sequence in \mathbb{R}^N , $\mathcal{L}(Y) = p$;

- $(I_k, k = 1, 2, \dots, n)$: a partition of $[1, N]$, $|I_k| = n_k$;
- for $x \in \mathbb{R}^N$, $x^{(k)} \triangleq \{x_i : i \in I_k\}$, $\bar{x}^{(k)} \triangleq \{x_i : i \notin I_k\}$;
- $X^{(k)}$ and $\bar{X}^{(k)}$: the corresponding segments of X ;
- $Y^{(k)}$ and $\bar{Y}^{(k)}$: the corresponding segments of Y ;
- $\bar{q}^{(k)} \triangleq \mathcal{L}(\bar{X}^{(k)})$, $Q^{(k)}(\cdot|\bar{x}^{(k)}) \triangleq \mathcal{L}(X^{(k)}|\bar{X}^{(k)} = \bar{x}^{(k)})$;
- $\bar{p}^{(k)} \triangleq \mathcal{L}(\bar{Y}^{(k)})$, $p^{(k)}(\cdot|\bar{y}^{(k)}) \triangleq \mathcal{L}(Y^{(k)}|\bar{Y}^{(k)} = \bar{y}^{(k)})$.

We consider \mathbb{R}^N as the product of Euclidean spaces $\mathbb{R}^{(k)}$ of dimension n_k .

Definition. The conditional distributions $Q^{(k)}(\cdot|\bar{x}^{(k)})$ and $p^{(k)}(\cdot|\bar{x}^{(k)})$ are called the local specifications of q resp. p .

To formulate the main results of this paper, we also need the concept of (average) conditional relative entropy, together with some more notation:

Definition. If we are given a probability measure $\pi = \mathcal{L}(S)$ on \mathbb{R}^ℓ ($\ell \geq 1$ integer), and conditional distributions $\mu(\cdot|s) = \mathcal{L}(Z|S = s)$, $\nu(\cdot|s) = \mathcal{L}(U|S = s)$ on \mathbb{R}^m then consider the average relative entropy

$$\mathbb{E}_\pi D(\mu(\cdot|S)\|\nu(\cdot|S)) = \int_{\mathbb{R}^\ell} D(\mu(\cdot|s)\|\nu(\cdot|s))\pi(ds).$$

For $\mathbb{E}_\pi D(\mu(\cdot|S)\|\nu(\cdot|S))$ we shall use either of the notations

$$D(\mu(\cdot|S)\|\nu(\cdot|S)), \quad D(\mu(\cdot|S)\|U|S), \quad D(Z|S)\|\nu(\cdot|S), \quad D(Z|S)\|U|S).$$

For a fixed measure q on \mathbb{R}^N , we want to derive an inequality of the form

$$D(p\|q) \leq \frac{1}{\rho} \cdot \sum_{k=1}^n \rho_k \cdot D(p^{(k)}(\cdot|\bar{Y}^{(k)})\|Q^{(k)}(\cdot|\bar{Y}^{(k)})) \quad \text{for all } p \text{ on } \mathbb{R}^N, \quad (2.1)$$

for some positive constants $\rho_k, 1 \leq k \leq n$, and ρ . I.e., we want to bound $D(p\|q)$ by a weighted sum of the “single phase” conditional entropies $D(p^{(k)}(\cdot|\bar{Y}^{(k)})\|Q^{(k)}(\cdot|\bar{Y}^{(k)}))$. A bound of type (2.1) holds only for a restricted class of probability measures q , and we want a sufficient condition for (2.1). Since relative entropy measures in a way how different probability measures are, inequality (2.1) allows us to conclude to closeness of p and q from the closeness of their local specifications. Moreover, an inequality of type (2.1) ensures that upper bounds for the “single phase” relative entropies $D(p^{(k)}(\cdot|\bar{y}^{(k)})\|Q^{(k)}(\cdot|\bar{y}^{(k)}))$ that hold uniformly in $\bar{y}^{(k)}$, yield a bound for $D(p\|q)$. This is a way to get logarithmic Sobolev inequalities for measures on product spaces.

To get inequality (2.1), we make three assumptions explained below. Recall that $(I_k, k = 1, 2, \dots, n)$ is a partition of $[1, N]$.

Assumption 1. Assume that $Q^{(k)}(\cdot|\bar{x}^{(k)})$ satisfies a logarithmic Sobolev inequality with constant ρ_k for all $x \in \mathbb{R}^N$ and $k \in [1, n]$.

Consider the Hessian of $V(x) = -\log q(x)$, i.e., the matrix $(V_{i,j}(x))_{i,j \in [1, N]}$, where we denote by $V_{i,j}(x)$ the second partial derivatives of $V(x)$.

Assumption 2. Assume that, for each $k \in [1, n]$, the matrix $(V_{i,j}(x))_{i,j \in I_k}$ is bounded from below by some (possibly negative) constant times the identity.

To formulate Assumption 3, we introduce the following

Notation. Under Assumption 1, and for sequences $x, \xi \in \mathbb{R}^N$ fixed, we denote by $A(x, \xi)$ the matrix with elements

$$\begin{aligned} A_{i,j}(x, \xi) &= \frac{V_{i,j}(\bar{x}^{(\ell)}, \xi^{(\ell)})}{\sqrt{\rho_k} \cdot \sqrt{\rho_\ell}} \quad \text{for } i \in I_k, j \in I_\ell, k \neq \ell, \\ A_{i,j}(x, \xi) &= 0 \quad \text{if } i \text{ and } j \text{ belong to the same set } I_k. \end{aligned}$$

Moreover, for sequences $x, \xi \in \mathbb{R}^N$ and $0 < \rho < \min \rho_k$, we denote by $A^\rho(x, \xi)$ the matrix with elements

$$\begin{aligned} A_{i,j}^\rho(x, \xi) &= \frac{V_{i,j}(\bar{x}^{(\ell)}, \xi^{(\ell)})}{\sqrt{\rho_k - \rho} \cdot \sqrt{\rho_\ell - \rho}} \quad \text{for } i \in I_k, j \in I_\ell, k \neq \ell, \\ A_{i,j}^\rho(x, \xi) &= 0 \quad \text{if } i \text{ and } j \text{ belong to the same set } I_k. \end{aligned}$$

(Thus $A(x, \xi) = A^0(x, \xi)$.)

Remark. Unless the matrix $A^\rho(x, \xi)$ is constant in x , it is not symmetric, since in the definition of $A_{i,j}^\rho(x, \xi)$ ($i \in I_k, j \in I_\ell$), we use $\xi^{(\ell)}$, and not $\xi^{(k)}$.

Assumption 3. We assume that

$$\sup_{x, \xi} \|A(x, \xi)\| \triangleq 1 - \delta < 1, \quad (2.2)$$

and that ρ is such that

$$\sup_{x, \xi} \|A^\rho(x, \xi)\| \leq 1. \quad (2.3)$$

Conditions (2.2) and (2.3) shall be used in the following form: For all $x, \xi, u, v \in \mathbb{R}^N$,

$$\begin{aligned} & \left| \sum_{k, \ell \in [1, n], k \neq \ell} \sum_{i \in I_k, j \in I_\ell} u_i \cdot V_{i,j}(x, \xi) \cdot v_j \right| \\ & \leq (1 - \delta) \cdot \sqrt{\sum_{k=1}^n \rho_k \cdot |u^{(k)}|^2} \cdot \sqrt{\sum_{\ell=1}^n \rho_\ell \cdot |v^{(\ell)}|^2} \end{aligned} \quad (2.4)$$

and

$$\begin{aligned} & \left| \sum_{k, \ell \in [1, n], k \neq \ell} \sum_{i \in I_k, j \in I_\ell} u_i \cdot V_{i,j}(x, \xi) \cdot v_j \right| \\ & \leq \sqrt{\sum_{k=1}^n (\rho_k - \rho) \cdot |u^{(k)}|^2} \cdot \sqrt{\sum_{\ell=1}^n (\rho_\ell - \rho) \cdot |v^{(\ell)}|^2}, \end{aligned} \quad (2.5)$$

respectively.

Theorem 1. *If Assumptions 1-3 hold then*

$$D(p||q) \leq \frac{1}{\rho} \cdot \sum_{k=1}^n \rho_k \cdot D\left(p^{(k)}(\cdot|\bar{Y}^{(k)}) || Q^{(k)}(\cdot|\bar{Y}^{(k)})\right) \quad (2.6)$$

for any probability measure p on \mathbb{R}^N .

Theorem 2. *Under Assumptions 1-3, q satisfies a logarithmic Sobolev inequality with constant ρ .*

Theorem 2 follows from Theorem 1, using Assumption 1 and the fact that by the definition of the operator ∇

$$I(p||q) = \sum_{k=1}^n \mathbb{E} I(p^{(k)}(\cdot|\bar{Y}^{(k)}) || Q^{(k)}(\cdot|\bar{Y}^{(k)})).$$

The statement of Theorem 2 was proved by F. Otto and M. Reznikoff [10], under a condition similar to, but stronger than, Assumption 3. We discuss Otto and Reznikoff's theorem at the end of this section.

Next we formulate a logarithmic Sobolev inequality for a discrete time Markov process governed by the local specifications $Q^{(k)}(\cdot|\bar{y}^{(k)})$.

Definition of weighted Gibbs sampler.

Given a partition $(I_k, k = 1, 2, \dots, n)$ of $[1, N]$, and local specifications $Q^{(k)}(\cdot|\bar{y}^{(k)})$, the weighted Gibbs sampler Γ with weights $(\pi^{(k)}, k = 1, 2, \dots, n)$ is the Markov operator on the probability measures p (on \mathbb{R}^N) defined by

$$\Gamma = \sum_{k=1}^n \pi_k \Gamma_k, \quad \Gamma_k(z|y) = \delta(\bar{y}^{(k)}, \bar{z}^{(k)}) \cdot Q^{(k)}(z^{(k)}|\bar{y}^{(k)}).$$

(Here δ denotes Kronecker's δ .)

Corollary to Theorem 1.

If Assumptions 1-3 hold then for the weighted Gibbs sampler Γ with weights

$$(\rho_k/R, k = 1, 2, \dots, n), \quad R = \sum_k \rho_k,$$

we have

$$D(p||q) \leq \frac{R}{\rho} \cdot \left(D(p||q) - D(p\Gamma||q) \right). \quad (2.7)$$

Thus

$$D(p\Gamma^m||q) \leq \left(1 - \frac{\rho}{R} \right)^m \cdot D(p||q).$$

(2.7) follows from Theorem 1 by the inequality

$$D(p\Gamma||q) \leq \frac{1}{R} \sum_{k=1}^n \rho_k D(p\Gamma_k||q)$$

(a consequence of the convexity of relative entropy) and the identity

$$D(p||q) - D(p\Gamma_k||q) = D\left(p^{(k)}(\cdot|\bar{Y}^{(k)}) || Q^{(k)}(\cdot|\bar{Y}^{(k)}) \right).$$

(2.7) can be considered as a logarithmic Sobolev inequality for the Gibbs sampler Γ . Indeed, for the Markov process defined by Γ , it bounds relative entropy (from the stationary distribution) by the decrease of relative entropy along the Markov process.

Next we formulate a transportation-cost inequality that follows from Theorem 2, using the Otto-Villani theorem (Theorem 1 in [3]). We need the following definitions:

Definition. The quadratic Wasserstein distance between the probability measures r and s on \mathbb{R}^m is defined as

$$W(r, s) = \inf_{\pi} [E_{\pi} |\xi - \eta|^2]^{1/2},$$

where ξ and η are random variables with laws r resp. s , $|\xi - \eta|$ denotes Euclidean distance, and infimum is taken over all distributions $\pi = \mathcal{L}(\xi, \eta)$ with marginals r and s .

Definition. A probability measure s on \mathbb{R}^m satisfies a transportation-cost inequality with constant ρ if

$$W^2(r, s) \leq \frac{2}{\rho} \cdot D(r \| s)$$

for all probability measures r on \mathbb{R}^m .

Transportation-cost inequalities are useful in proving measure concentration inequalities. A transportation-cost inequality for the case when q is Gaussian, was proved by Talagrand [11]. Otto and Villani generalized Talagrand's inequality as follows:

Otto and Villani's theorem for Euclidean spaces. [3],[12]

If a density function on \mathbb{R}^m satisfies a logarithmic Sobolev inequality then it satisfies a transportation-cost inequality with the same constant.

By Otto and Villani's theorem, Theorem 2 implies the following

Theorem 3.

If Assumptions 1-3 hold then q satisfies a transportation-cost inequality with constant ρ .

In [13], corrected in [14], the statement of Theorem 3, for equal ρ_k 's, was proved modulo an absolute constant factor.

Now we compare Theorem 2 with the result of [10].

In [10] the statement of Theorem 2 is proved under the following condition in place of (2.3):

For $k, \ell \in [1, n]$, $k \neq \ell$, and $x \in \mathbb{R}^N$, consider the following minors of the Hessian of $V(x)$:

$$K_{k,\ell}(x) = \left(V_{i,j}(x) \right)_{i \in I_k, j \in I_\ell},$$

and set

$$\kappa_{k,\ell} = \sup_x \left\| (K_{k,\ell}(x)) \right\|.$$

Then consider the $n \times n$ matrix

$$K = (\kappa_{k,\ell})_{k,\ell \in [1,n], k \neq \ell}.$$

(K has 0's in the main diagonal.) Otto and Reznikoff use the assumption that

$$K \leq \Lambda(\{\rho_k - \rho\}), \tag{2.3'}$$

where $\Lambda(\{\rho_k - \rho\})$ denotes the $n \times n$ diagonal matrix with elements $\rho_k - \rho$. With the notation

$$\kappa'_{k,\ell} = \frac{\kappa_{k,\ell}}{\sqrt{\rho_k - \rho} \cdot \sqrt{\rho_\ell - \rho}}$$

$$K'^\rho = (\kappa'_{k,\ell})_{k,\ell \in [1,n], k \neq \ell},$$

(2.3') can be written in the form

$$K'^\rho \leq Id,$$

where Id is the $n \times n$ identity matrix. Since K'^ρ is symmetric, this means that the largest eigenvalue of K'^ρ is ≤ 1 . The elements of K'^ρ are non-negative, thus, by Perron's theorem, the largest eigenvalue of K'^ρ equals $\|K'^\rho\|$. I.e., in [10] it is actually assumed that

$$\|K'^\rho\| \leq 1, \quad (2.3'')$$

which is clearly stronger than (2.3).

Remark.

If q is Gaussian then the Hessian of $V(x)$ does not depend on x . Otto and Reznikoff's result is tight for Gaussian distributions with attractive interactions. (For $\mathbb{R}^{(k)} = \mathbb{R}$; attractivity means that $V_{i,j} \leq 0$ for $i \neq j$.) For q Gaussian and $\mathbb{R}^{(k)} = \mathbb{R}$, Theorem 2 can be formulated as follows: If $\|A^0\| < 1$ then q satisfies a logarithmic Sobolev inequality with constant ρ , where ρ is the largest number satisfying

$$\|A^\rho\| = 1. \quad (2.8)$$

Thus Theorem 2 is tight for those Gaussian distributions q for which $\|A^\rho\|$ (for the ρ defined by (2.8)) is given by the absolute value of the smallest negative eigenvalue (and not the largest positive one).

Example.

Assumption 3 is practically impossible to check, except when the mixed partial derivatives of $V(x)$ are constants. Otherwise we probably cannot do better than use Otto and Reznikoff's theorem. However, if the mixed partial derivatives of $V(x)$ are all constants then Theorem 2 may give a better result. Indeed, let $V(x) = -\log q(x)$ be of the form

$$V(x) = \sum_{k=1}^n \phi_k(x) + \sum_{k,\ell \in [1,n], k \neq \ell} a_{k,\ell} \cdot x_k \cdot x_\ell,$$

where for each k and fixed \bar{x}_k , the single phase density $C_k(\bar{x}_k) \cdot \exp(-\phi_k(x_k, \bar{x}_k))$, as a function of x_k , satisfies a logarithmic Sobolev inequality with a common constant ρ . Theorem 2 guaranties a positive logarithmic Sobolev constant if the matrix with elements

$$a_{k,\ell} \quad \text{outside the main diagonal, and} \quad 0 \quad \text{otherwise}$$

has norm $< \rho$. On the other hand, Otto and Reznikoff's theorem guaranties a positive logarithmic Sobolev constant if the matrix with elements

$$|a_{k,\ell}| \quad \text{outside the main diagonal, and} \quad 0 \quad \text{otherwise}$$

has norm $< \rho$. To see a concrete example when the first condition holds, but the second does not, consider the infinite dimensional Toeplitz matrix $B = (b_{k,\ell})$ defined by

$$b_{k,k+1} = 1, \quad b_{k,k+2} = -1, \quad b_{k,\ell} = 0 \quad \text{for} \quad \ell \geq k, \quad \ell \notin \{k+1, k+2\},$$

and $b_{k,\ell} = b_{\ell,k}$. From the theory of Toeplitz matrices (c.f. [15]) we know that

$$\| B \| = 2 \cdot \max |\cos x - \cos(2x)| = \frac{9}{4},$$

while for the matrix $abs(B)$ consisting of the absolute values of $b_{k,\ell}$, we get

$$\| abs(B) \| = 2 \cdot \max |\cos x + \cos(2x)| = 4.$$

Denote by B_m and $abs(B_m)$ the matrices consisting of the first m rows and columns of B resp. $abs(B)$; clearly $\|B_m\| \leq 9/4$ and $\lim_{m \rightarrow \infty} \|abs(B_m)\| = 4$. Therefore, if we take $A = B_m$, and if the functions $\phi_k(x)$ in the definition of $V(x)$ are such that the single phase densities $C_k(\bar{x}_k) \cdot \exp(-\phi_k(x_k, \bar{x}_k))$ satisfy a logarithmic Sobolev inequality with a common constant $> \frac{9}{4}$ then Theorem 2 guaranties a positive logarithmic Sobolev constant for $q = \exp(-V)$. However, we cannot get this from Otto and Reznikoff's theorem.

3. Proof of Theorem 1.

Our approach to prove Theorem 1 is based on the interpolation between the probability measures p and q realized by the solution of the Fokker-Planck equation

$$\partial_t p_t(y) = \Delta p_t(y) + \nabla \cdot (p_t(y) \cdot \nabla V(y)), \quad p_0(y) = p(y). \quad (3.1)$$

With the notation

$$h = p/q \quad \text{and} \quad h_t = p_t/q,$$

the Fokker-Planck equation (3.1) can be rewritten as follows:

$$\partial_t h_t = Lh_t \triangleq \Delta h_t - \nabla h_t \cdot \nabla V, \quad h_0(y) = h(y). \quad (3.2)$$

We have

$$\mathbb{E}_p \log h = D(p||q) \quad \text{and} \quad \mathbb{E}_{p_t} \log h_t = D(p_t||q).$$

Our argument heavily draws on the ideas developed in the paper by F. Otto and C. Villani [3]. To be able to use the tools of [3], we need the limit relation

$$\lim_{t \rightarrow \infty} D(p_t || q) = 0. \quad (3.3)$$

To this end we prove a logarithmic Sobolev inequality for q with a much smaller constant than claimed in Theorem 2. (It is disturbing that this weak preliminary result requires a very lengthy proof.)

Auxiliary Theorem. *If Assumptions 1-3 hold then q satisfies a logarithmic Sobolev inequality with a constant $C = C(R, \rho_{\min}, \delta)$, where $R = \sum_{k=1}^n \rho_k$, $\rho_{\min} = \min_k \rho_k$ and $\delta = 1 - \sup_{x, \xi} ||A(x, \xi)||$.*

For the proof of Theorem 1 we also need the following simple lemma (c. f. (32) in [3]).

Approximation Lemma. *In the proof of Theorem 1 we can restrict ourselves to the case when $V(x) = -\log q(x) \in C^\infty$, and $h(x) = p(x)/q(x)$ is of the form*

$$h(x) = (1 - \varepsilon) \cdot g(x) + \varepsilon, \quad \text{where} \\ g \in C^\infty \quad \text{is a compactly supported density function (with respect to } q), \quad \text{and} \quad \varepsilon > 0.$$

The proofs of the Auxiliary Theorem and the Approximation Lemma are postponed to Section 4, although they are used in the proof of Theorem 1 in this section.

We need some more

Notation.

Let

$$Y_t = (Y_{t,1}, Y_{t,2}, \dots, Y_{t,N})$$

denote a random sequence with $\mathcal{L}(Y_t) = p_t$, where p_t is the solution of the Fokker-Planck equation (3.1). In accordance with the notation at the beginning of Section 2, we write

$$Y_t^{(k)} = \{Y_{t,i} : i \in I_k\}, \quad \bar{Y}_t^{(k)} = \{Y_{t,i} : i \notin I_k\}.$$

Further, we set

$$\bar{p}_t^{(k)} = \mathcal{L}(\bar{Y}_t^{(k)}), \quad p_t^{(k)}(\cdot | \bar{y}_t^{(k)}) = \mathcal{L}(Y_t^{(k)} | \bar{Y}_t^{(k)} = \bar{y}_t^{(k)}).$$

By the Approximation Lemma we may assume that $V \in \mathcal{C}^\infty$. Then the domain of the operator L in (3.2) can be defined so as to contain the class \mathcal{D}_0 of those functions h in \mathcal{C}^∞ that are bounded, and whose partial derivatives of any order, multiplied by the partial derivatives of V of any order, are bounded. The class \mathcal{D}_0 is dense in $\mathbb{L}_2(q)$ and stable under L .

Again by the Approximation Lemma we can assume that the function $h_0 = h$ in (3.2) belongs to \mathcal{D}_0 . As explained in [3], this implies that h_t is uniformly bounded from below and from above, and that, for t fixed, $|\nabla h_t|^2$ is bounded. (Here we use the fact that, by Assumptions 2 and 3, the Hessian of $V(x)$ is bounded from below by a (possibly negative) constant times the identity.— In [3] assumption (32) of that paper is used which is implied by the assumption $h_0 = h \in \mathcal{D}_0$.)

Consequently, as explained in [3], under condition $h_0 = h \in \mathcal{D}_0$, the Fokker-Planck equation (3.2) defines a semigroup of diffeomorphisms

$$\Phi_t : \mathbb{R}^N \mapsto \mathbb{R}^N, \quad 0 \leq t < \infty, \quad (3.4)$$

satisfying

$$\partial_t \Phi_t(y) = -\nabla \log h_t(\Phi_t(y)), \quad (3.5)$$

and

$$p_t \Phi_s = p_{t+s}. \quad (3.6)$$

(3.6) means that p_{t+s} is the image of p_t under the map Φ_s . Since $\mathcal{L}(Y_t) = p_t$, we can think of the random sequences Y_t as functions of $Y = Y_0$:

$$Y_t = \Phi_t(Y) = \Phi_t(Y_0).$$

Let us introduce the function

$$\chi(y) = \sum_{k=1}^n \rho_k [\log h(y) - \log \bar{h}^{(k)}(\bar{y}^{(k)})], \quad y \in \mathbb{R}^N,$$

where

$$\bar{h}^{(k)}(\bar{y}^{(k)}) = \frac{\bar{p}^{(k)}(\bar{y}^{(k)})}{\bar{q}^{(k)}(\bar{y}^{(k)})} = \int_{\mathbb{R}^{(k)}} h(y) Q^{(k)}(dy^{(k)}) | \bar{y}^{(k)}.$$

(The integration domain is \mathbb{R}^{n_k} ; the superscript (k) indicates that integration is with respect to the variable $y^{(k)}$.) We have

$$\mathbb{E}_p \chi = \sum_{k=1}^n \rho_k \cdot D\left(p^{(k)}(\cdot | \bar{Y}^{(k)}) \parallel Q^{(k)}(\cdot | \bar{Y}^{(k)})\right).$$

Thus the statement of Theorem 1 is equivalent to

$$\rho \cdot \mathbb{E}_p \log h \leq \mathbb{E}_p \chi.$$

It is well known (and a proof can be found in [3]) that

$$\frac{\partial}{\partial t} D(p_t || q) = -I(p_t || q) = -\mathbb{E}_{p_t} |\nabla \log h_t|^2.$$

Thus, by (3.3),

$$D(p || q) = D(p || q) - \lim_{t \rightarrow \infty} D(p_t || q) = \int_0^\infty \mathbb{E}_{p_t} |\nabla \log h_t|^2 dt. \quad (3.7)$$

We introduce, analogously to the definition of χ , the functions

$$\chi_t(y) = \sum_{k=1}^n \rho_k [\log h_t(y) - \log \bar{h}_t^{(k)}(\bar{y}^{(k)})],$$

where

$$\bar{h}_t^{(k)}(\bar{y}^{(k)}) = \bar{p}_t^{(k)}(\bar{y}^{(k)}) / \bar{q}^{(k)}(\bar{y}^{(k)}).$$

We have

$$\mathbb{E}_{p_t} \chi_t = \sum_{k=1}^n \rho_k \cdot D\left(p_t^{(k)}(\cdot | \bar{Y}_t^{(k)}) \parallel Q^{(k)}(\cdot | \bar{Y}_t^{(k)})\right).$$

In particular, $\mathbb{E}_{p_t} \chi_t \geq 0$.

Using (3.7) and the fact that $\mathbb{E}_{p_t} \chi_t \geq 0$, for the proof of Theorem 1 it is enough to prove the following two propositions:

Proposition 1.

$$\mathbb{E}_p \chi - \lim_{t \rightarrow \infty} \mathbb{E}_{p_t} \chi_t = \int_0^\infty \mathbb{E}_{p_t} \left\{ \nabla \chi_t \cdot \nabla \log h_t \right\} dt. \quad (3.8)$$

Proposition 2.

$$\mathbb{E}_{p_t} \left\{ \nabla \chi_t \cdot \nabla \log h_t \right\} \geq \rho \cdot \mathbb{E}_{p_t} |\nabla \log h_t|^2.$$

Proof of Proposition 1.

For all $y \in \mathbb{R}^N$ we have

$$\chi(y) - \lim_{t \rightarrow \infty} \chi_t(\Phi_t(y)) = - \int_0^\infty \frac{\partial}{\partial t} (\chi_t(\Phi_t(y))) dt.$$

Therefore, by Fubini's theorem,

$$\mathbb{E}_p \left\{ \chi(Y) - \lim_{t \rightarrow \infty} \chi_t(\Phi_t(Y)) \right\} = - \int_0^\infty \mathbb{E}_p \left\{ \frac{\partial}{\partial t} (\chi_t(\Phi_t(Y))) \right\} dt. \quad (3.9)$$

Denoting by dot derivation with respect to t , and using (3.5):

$$\frac{\partial}{\partial t} \left(\chi_t(\Phi_t(y)) \right) = \dot{\chi}_t(\Phi_t(y)) - \nabla \chi_t(\Phi_t(y)) \cdot \nabla \log h_t(\Phi_t(y)). \quad (3.10)$$

Further,

$$\partial_t (\chi_t(z)) = \sum_{k=1}^n \rho_k \cdot \left[\partial_t \left(\log h_t(z) \right) - \partial_t \left(\log \bar{h}_t^{(k)}(\bar{z}^{(k)}) \right) \right], \quad z \in \mathbb{R}^N. \quad (3.11)$$

To calculate $\partial_t (\log \bar{h}_t^{(k)}(\bar{z}^{(k)}))$, we need the following

Lemma.

The solution h_t of the Fokker-Planck equation (3.2) satisfies

$$\|\partial_t h_t\|_{\mathbb{L}_2(q)} \leq \|L h_0\|_{\mathbb{L}_2(q)} < \infty. \quad (3.12)$$

Proof.

The operator L is defined on a dense subset \mathcal{D}_0 of $\mathbb{L}_2(q)$. Moreover, L is symmetric and negative definite on \mathcal{D}_0 . Indeed, by partial integration we have

$$(Lf, g)_{L_2(q)} = \int_{\mathbb{R}^N} (\Delta f - \nabla V \cdot \nabla f) \cdot g dq = - \int_{\mathbb{R}^N} \nabla f \cdot \nabla g dq.$$

It follows that for $\lambda > 0$

$$((\lambda I - L)f, f)_{\mathbb{L}_2(q)} \geq \lambda \cdot \|f\|_{\mathbb{L}_2(q)}^2,$$

i.e.,

$$\|(\lambda I - L)^{-1}\| \leq \frac{1}{\lambda}.$$

Thus by the Hille-Yosida theorem (c.f. [16]), there exists a contraction semigroup $(P_t : t \geq 0)$ on $\mathbb{L}_2(q)$ whose generator is L :

$$\partial_t P_t h_0 = L P_t h_0 \quad \text{for } h_0 \in \mathcal{D}_0, \quad \text{and} \quad \|P_t\| \leq 1.$$

For $h_0 \in \mathcal{D}_0$, the solution of (3.2) can be written as $h_t = P_t h_0$, and since $P_t L = L P_t$, we have

$$\partial_t h_t = \partial_t P_t h_0 = L P_t h_0 = P_t L h_0,$$

which implies (3.12). \square

By the above Lemma, $\partial_t h_t \in \mathbb{L}_1(q)$, so we can differentiate under the integral sign in the next formula:

$$\begin{aligned} \partial_t \left(\log \bar{h}_t^{(k)}(\bar{z}^{(k)}) \right) &= \partial_t \int_{\mathbb{R}^{(k)}} h_t(z) Q^{(k)}(dz^{(k)} | \bar{z}^{(k)}) \\ &= \frac{\int_{\mathbb{R}^{(k)}} \partial_t \left(h_t(z) \right) Q^{(k)}(dz^{(k)} | \bar{z}^{(k)})}{\bar{h}_t^{(k)}(\bar{z}^{(k)})} \\ &= \int_{\mathbb{R}^{(k)}} \partial_t \log h_t(z) \cdot \frac{h_t(z)}{\bar{h}_t^{(k)}(\bar{z}^{(k)})} Q^{(k)}(dz^{(k)} | \bar{z}^{(k)}). \end{aligned} \quad (3.13)$$

By the definition of the function h_t ,

$$\frac{h_t(z)}{\bar{h}_t^{(k)}(\bar{z}^{(k)})} Q^{(k)}(dz^{(k)} | \bar{z}^{(k)}) = p_t^{(k)}(dz^{(k)} | \bar{z}^{(k)}).$$

Thus (3.13) implies

$$\begin{aligned} \partial_t \left(\log \bar{h}_t^{(k)}(\bar{z}^{(k)}) \right) &= \int_{\mathbb{R}^{(k)}} \partial_t \log h_t(z) p_t^{(k)}(dz^{(k)} | \bar{z}^{(k)}) = \mathbb{E}_{p_t} \{ \partial_t \log h_t | \bar{z}^{(k)} \}, \end{aligned} \quad (3.14)$$

where $\bar{z}^{(k)}$ in the condition of the expectation is a shorthand for $\bar{Y}_t^{(k)} = \bar{z}^{(k)}$. Substituting (3.14) into (3.11) we get

$$\partial_t (\chi_t(z)) = \sum_{k=1}^n \rho_k \cdot \left[\partial_t \log h_t(z) - \mathbb{E}_{p_t} \{ \partial_t \log h_t | \bar{z}^{(k)} \} \right].$$

It follows that $\mathbb{E}_{p_t} \dot{\chi}_t = 0$ which, together with (3.10), yields

$$\mathbb{E}_p \frac{\partial}{\partial t} \left(\chi_t(\Phi_t(Y)) \right) = -\mathbb{E}_{p_t} \left\{ \nabla \chi_t \cdot \nabla \log h_t \right\}.$$

Substituting this into (3.9) we get (3.8). □

Proof of Proposition 2.

We prove Proposition 2 for $t = 0$; for $t > 0$ the proof is the same. For a function $g : \mathbb{R}^N \mapsto \mathbb{R}$ set

$$\nabla^{(k)} g(x) = (\partial_i g(x) : i \in I_k).$$

We need the following

Proposition 3.

For $k, \ell \in [1, n]$, $k \neq \ell$, we have

$$\begin{aligned} \nabla^{(k)} \log \bar{h}^{(\ell)}(\bar{y}^{(\ell)}) &= \mathbb{E}_p \{ \nabla^{(k)} \log h | \bar{y}^{(\ell)} \} \\ &- \int_{\mathbb{R}^{(\ell)} \times \mathbb{R}^{(\ell)}} [\nabla^{(k)} V(\bar{y}^{(\ell)}, \xi^{(\ell)}) - \nabla^{(k)} V(\bar{y}^{(\ell)}, \eta^{(\ell)})] \pi^{(\ell)}(d\xi^{(\ell)}, d\eta^{(\ell)} | \bar{y}^{(\ell)}), \end{aligned} \quad (3.15)$$

where $\pi^{(\ell)}(d\xi^{(\ell)}, d\eta^{(\ell)} | \bar{y}^{(\ell)})$ is an arbitrary coupling of the conditional measures $p^{(\ell)}(\cdot | \bar{y}^{(\ell)})$ and $Q^{(\ell)}(\cdot | \bar{y}^{(\ell)})$. (I.e., $\pi^{(\ell)}(d\xi^{(\ell)}, d\eta^{(\ell)} | \bar{y}^{(\ell)})$ is a conditional density on $\mathbb{R}^{(\ell)} \times \mathbb{R}^{(\ell)}$ with marginals $p^{(\ell)}(\cdot | \bar{y}^{(\ell)})$ and $Q^{(\ell)}(\cdot | \bar{y}^{(\ell)})$.)

Proof of Proposition 3.

Since $|\nabla h|$ is bounded (and $|\nabla h_t|$ is also bounded for t fixed), we have

$$\nabla^{(k)} \bar{h}^{(\ell)}(\bar{y}^{(\ell)}) = \int_{\mathbb{R}^{(\ell)}} \nabla^{(k)} \left(h(\bar{y}^{(\ell)}, \xi^{(\ell)}) \cdot Q^{(\ell)}(\xi^{(\ell)} | \bar{y}^{(\ell)}) \right) d\xi^{(\ell)}. \quad (3.16)$$

Further,

$$\begin{aligned} \nabla^{(k)} Q^{(\ell)}(\xi^{(\ell)} | \bar{y}^{(\ell)}) &= \nabla^{(k)} \frac{\exp(-V(\bar{y}^{(\ell)}, \xi^{(\ell)}))}{\int_{\mathbb{R}^{(\ell)}} \exp(-V(\bar{y}^{(\ell)}, \eta^{(\ell)})) d\eta^{(\ell)}} \\ &= -\nabla^{(k)} V(\bar{y}^{(\ell)}, \xi^{(\ell)}) \cdot Q^{(\ell)}(\xi^{(\ell)} | \bar{y}^{(\ell)}) \\ &\quad + Q^{(\ell)}(\xi^{(\ell)} | \bar{y}^{(\ell)}) \cdot \int_{\mathbb{R}^{(\ell)}} \nabla^{(k)} V(\bar{y}^{(\ell)}, \eta^{(\ell)}) Q^{(\ell)}(d\eta^{(\ell)} | \bar{y}^{(\ell)}) \\ &= Q^{(\ell)}(\xi^{(\ell)} | \bar{y}^{(\ell)}) \cdot \int_{\mathbb{R}^{(\ell)}} \left[\nabla^{(k)} V(\bar{y}^{(\ell)}, \eta^{(\ell)}) - \nabla^{(k)} V(\bar{y}^{(\ell)}, \xi^{(\ell)}) \right] Q^{(\ell)}(d\eta^{(\ell)} | \bar{y}^{(\ell)}). \end{aligned}$$

It follows that

$$\begin{aligned}
& \nabla^{(k)} \left(h(\bar{y}^{(\ell)}, \xi^{(\ell)}) \cdot Q^{(\ell)}(\xi^{(\ell)} | \bar{y}^{(\ell)}) \right) \\
&= Q^{(\ell)}(\xi^{(\ell)} | \bar{y}^{(\ell)}) \cdot \left[\nabla^{(k)} h(\bar{y}^{(\ell)}, \xi^{(\ell)}) \right. \\
&\quad \left. + h(\bar{y}^{(\ell)}, \xi^{(\ell)}) \cdot \int_{\mathbb{R}^{(\ell)}} \left(\nabla^{(k)} V(\bar{y}^{(\ell)}, \eta^{(\ell)}) - \nabla^{(k)} V(\bar{y}^{(\ell)}, \xi^{(\ell)}) \right) Q^{(\ell)}(d\eta^{(\ell)} | \bar{y}^{(\ell)}) \right].
\end{aligned} \tag{3.17}$$

Substituting (3.17) into (3.16):

$$\begin{aligned}
& \nabla^{(k)} \bar{h}^{(\ell)}(\bar{y}^{(\ell)}) \\
&= \int_{\mathbb{R}^{(\ell)}} \nabla^{(k)} h(\bar{y}^{(\ell)}, \xi^{(\ell)}) Q^{(\ell)}(d\xi^{(\ell)} | \bar{y}^{(\ell)}) \\
&\quad + \int_{\mathbb{R}^{(\ell)}} h(\bar{y}^{(\ell)}, \xi^{(\ell)}) Q^{(\ell)}(d\xi^{(\ell)} | \bar{y}^{(\ell)}) \cdot \int_{\mathbb{R}^{(\ell)}} \nabla^{(k)} V(\bar{y}^{(\ell)}, \eta^{(\ell)}) Q^{(\ell)}(d\eta^{(\ell)} | \bar{y}^{(\ell)}) \\
&\quad - \int_{\mathbb{R}^{(\ell)}} h(\bar{y}^{(\ell)}, \xi^{(\ell)}) \cdot \nabla^{(k)} V(\bar{y}^{(\ell)}, \xi^{(\ell)}) Q^{(\ell)}(d\xi^{(\ell)} | \bar{y}^{(\ell)}) \\
&= \int_{\mathbb{R}^{(\ell)}} \nabla^{(k)} h(\bar{y}^{(\ell)}, \xi^{(\ell)}) Q^{(\ell)}(d\xi^{(\ell)} | \bar{y}^{(\ell)}) + \bar{h}^{(\ell)}(\bar{y}^{(\ell)}) \cdot \int_{\mathbb{R}^{(\ell)}} \nabla^{(k)} V(\bar{y}^{(\ell)}, \eta^{(\ell)}) Q^{(\ell)}(d\eta^{(\ell)} | \bar{y}^{(\ell)}) \\
&\quad - \int_{\mathbb{R}^{(\ell)}} h(\bar{y}^{(\ell)}, \xi^{(\ell)}) \cdot \nabla^{(k)} V(\bar{y}^{(\ell)}, \xi^{(\ell)}) Q^{(\ell)}(d\xi^{(\ell)} | \bar{y}^{(\ell)}).
\end{aligned}$$

Dividing both sides by $\bar{h}^{(\ell)}(\bar{y}^{(\ell)})$:

$$\begin{aligned}
& \nabla^{(k)} \log \bar{h}^{(\ell)}(\bar{y}^{(\ell)}) = \int_{\mathbb{R}^{(\ell)}} \nabla^{(k)} \log h(\bar{y}^{(\ell)}, \xi^{(\ell)}) \cdot \frac{h(\bar{y}^{(\ell)}, \xi^{(\ell)})}{\bar{h}^{(\ell)}(\bar{y}^{(\ell)})} Q^{(\ell)}(d\xi^{(\ell)} | \bar{y}^{(\ell)}) \\
&\quad + \int_{\mathbb{R}^{(\ell)}} \nabla^{(k)} V(\bar{y}^{(\ell)}, \eta^{(\ell)}) Q^{(\ell)}(d\eta^{(\ell)} | \bar{y}^{(\ell)}) \\
&\quad - \int_{\mathbb{R}^{(\ell)}} \nabla^{(k)} V(\bar{y}^{(\ell)}, \xi^{(\ell)}) \cdot \frac{h(\bar{y}^{(\ell)}, \xi^{(\ell)})}{\bar{h}^{(\ell)}(\bar{y}^{(\ell)})} Q^{(\ell)}(d\xi^{(\ell)} | \bar{y}^{(\ell)}).
\end{aligned}$$

Since

$$\frac{h(\bar{y}^{(\ell)}, \xi^{(\ell)})}{\bar{h}^{(\ell)}(\bar{y}^{(\ell)})} Q^{(\ell)}(d\xi^{(\ell)} | \bar{y}^{(\ell)}) = p^{(\ell)}(d\xi^{(\ell)} | \bar{y}^{(\ell)}),$$

and

$$\int_{\mathbb{R}^{(\ell)}} \nabla^{(k)} \log h(\bar{y}^{(\ell)}, \xi^{(\ell)}) p^{(\ell)}(d\xi^{(\ell)} | \bar{y}^{(\ell)}) = \mathbb{E}_p \{ \nabla^{(k)} \log h | \bar{y}^{(\ell)} \},$$

(3.15) follows. □

Now we are ready to prove Proposition 2. By Proposition 3 we have

$$\begin{aligned}\nabla^{(k)}\chi(y) &= \sum_{\ell=1}^n \rho_\ell \cdot \left[\nabla^{(k)} \log h(y) - \nabla^{(k)} \log \bar{h}^{(\ell)}(y^{(\ell)}) \right] \\ &= \sum_{\ell=1}^n \rho_\ell \cdot \left[\nabla^{(k)} \log h(y) - \mathbb{E}_p \{ \nabla^{(k)} \log \bar{h} | \bar{y}^{(\ell)} \} \right] \\ &\quad + \int_{\mathbb{R}^N \times \mathbb{R}^N} \left(\sum_{\ell \neq k} \rho_\ell \cdot \left[\nabla^{(k)} V(\bar{y}^{(\ell)}, \xi^{(\ell)}) - \nabla^{(k)} V(\bar{y}^{(\ell)}, \eta^{(\ell)}) \right] \right) \Pi(d\xi, d\eta|y),\end{aligned}$$

where $\Pi(d\xi, d\eta|y)$ denotes the conditional product measure $\prod_{\ell=1}^n \pi^{(\ell)}(d\xi^{(\ell)}, d\eta^{(\ell)} | \bar{y}^{(\ell)})$.

It follows that

$$\begin{aligned}\mathbb{E}_p \{ \nabla \chi \cdot \nabla \log h \} &= \sum_{k=1}^n \rho_k \cdot \mathbb{E}_p | \nabla^{(k)} \log h |^2 \\ &\quad + \sum_{k, \ell \in [1, n], \ell \neq k} \rho_\ell \cdot \left[\mathbb{E}_p | \nabla^{(k)} \log h |^2 - \mathbb{E}_p \left\{ \mathbb{E}_p \{ \nabla^{(k)} \log h | \bar{y}^{(\ell)} \} \cdot \nabla^{(k)} \log h \right\} \right] \\ &\quad + \mathbb{E}_{p, \Pi} \left\{ \sum_{k, \ell \in [1, n], \ell \neq k} \rho_\ell \cdot \left[\nabla^{(k)} V(\bar{y}^{(\ell)}, \xi^{(\ell)}) - \nabla^{(k)} V(\bar{y}^{(\ell)}, \eta^{(\ell)}) \right] \cdot \nabla^{(k)} \log h \right\},\end{aligned}\tag{3.18}$$

Here $\mathbb{E}_{p, \Pi}$ denotes expectation with respect to the joint distribution $\mathcal{L}(Y, \xi, \eta)$, defined by $\mathcal{L}(Y) = p$ and $\mathcal{L}(\xi, \eta|Y) = \Pi(d\xi, d\eta|y)$.

For $k \neq \ell$ we have

$$\begin{aligned}\mathbb{E}_p \left\{ \left[\nabla^{(k)} \log h(y) - \mathbb{E}_p \{ \nabla^{(k)} \log h | \bar{y}^{(\ell)} \} \right] \cdot \nabla^{(k)} \log h(y) \right\} \\ = \mathbb{E}_p \left| \nabla^{(k)} \log h(y) \right|^2 - \mathbb{E}_p \left\{ \mathbb{E}_p^2 \{ \nabla^{(k)} \log h | \bar{y}^{(\ell)} \} \right\} \geq 0.\end{aligned}\tag{3.19}$$

To estimate the last line in (3.18), we introduce the notation

$$U(y, \xi) = \sum_{k, \ell \in [1, n], \ell \neq k} \rho_\ell \cdot \nabla^{(k)} V(\bar{y}^{(\ell)}, \xi^{(\ell)}) \cdot \nabla^{(k)} \log h(y), \quad y, \xi \in \mathbb{R}^N.$$

We have

$$\sum_{k, \ell \in [1, n], \ell \neq k} \rho_\ell \cdot \left[\nabla^{(k)} V(\bar{y}^{(\ell)}, \xi^{(\ell)}) - \nabla^{(k)} V(\bar{y}^{(\ell)}, \eta^{(\ell)}) \right] \cdot \nabla^{(k)} \log h(y) = U(y, \xi) - U(y, \eta).$$

To estimate $|U(y, \xi) - U(y, \eta)|$, we carry out the following calculation:

$$\begin{aligned}
& \frac{\partial}{\partial \tau} U(y, \eta + \tau(\xi - \eta)) \\
&= \sum_{k, \ell \in [1, n], \ell \neq k} \sum_{i \in I_k, j \in I_\ell} \rho_\ell \cdot (\xi_j - \eta_j) \cdot V_{i,j}(\bar{y}^{(\ell)}, \eta^{(\ell)} + \tau(\xi^{(\ell)} - \eta^{(\ell)})) \cdot \partial_i \log h(y).
\end{aligned}$$

Hence, by Assumption 3 (c.f. (2.5)),

$$\begin{aligned}
& \left| \frac{\partial}{\partial \tau} U(y, \eta + \tau(\xi - \eta)) \right| \\
& \leq \sqrt{\sum_{\ell=1}^n \sum_{j \in I_\ell} (\rho_\ell - \rho) \cdot \rho_\ell^2 \cdot (\xi_j - \eta_j)^2} \cdot \sqrt{\sum_{k=1}^n \sum_{i \in I_k} (\rho_k - \rho) \cdot |\partial_i \log h(y)|^2} \\
& = \sqrt{\sum_{\ell=1}^n (\rho_\ell - \rho) \cdot \rho_\ell^2 \cdot (\xi^{(\ell)} - \eta^{(\ell)})^2} \cdot \sqrt{\sum_{k=1}^n (\rho_k - \rho) \cdot |\nabla^{(k)} \log h(y)|^2}.
\end{aligned}$$

It follows that for all y, ξ and η

$$\begin{aligned}
& \left| \sum_{k, \ell \in [1, n], \ell \neq k} \rho_\ell \cdot \left[\nabla^{(k)} V(\bar{y}^{(\ell)}, \xi^{(\ell)}) - \nabla^{(k)} V(\bar{y}^{(\ell)}, \eta^{(\ell)}) \right] \cdot \nabla^{(k)} \log h(y) \right| \\
& = \left| U(y, \xi) - U(y, \eta) \right| \\
& \leq \sqrt{\sum_{\ell=1}^n (\rho_\ell - \rho) \cdot \rho_\ell^2 \cdot |\xi^{(\ell)} - \eta^{(\ell)}|^2} \cdot \sqrt{\sum_{k=1}^n (\rho_k - \rho) \cdot |\nabla^{(k)} \log h(y)|^2}.
\end{aligned}$$

Now the last line of (3.18) can be estimated as follows:

$$\begin{aligned}
& \mathbb{E}_{p, \Pi} \left| \left\{ \sum_{k, \ell \in [1, n], \ell \neq k} \rho_\ell \cdot \left[\nabla^{(k)} V(\bar{y}^{(\ell)}, \xi^{(\ell)}) - \nabla^{(k)} V(\bar{y}^{(\ell)}, \eta^{(\ell)}) \right] \cdot \nabla^{(k)} \log h \right\} \right| \\
& \leq \sqrt{\sum_{\ell=1}^n (\rho_\ell - \rho) \cdot \rho_\ell^2 \cdot \mathbb{E}_{p, \Pi} |\xi^{(\ell)} - \eta^{(\ell)}|^2} \cdot \sqrt{\sum_{k=1}^n (\rho_k - \rho) \cdot \mathbb{E}_p |\nabla^{(k)} \log h(y)|^2}. \tag{3.20}
\end{aligned}$$

Our calculations are valid for any coupling of the conditional densities $p^{(\ell)}(d\xi^{(\ell)}|\bar{y}^{(\ell)})$ and $Q^{(\ell)}(d\eta^{(\ell)}|\bar{y}^{(\ell)})$. Now we specify $\pi^{(\ell)}(d\xi^{(\ell)}, d\eta^{(\ell)}|\bar{y}^{(\ell)})$ so as to achieve

$$\mathbb{E}_{\pi_\ell} \{ |\eta^{(\ell)} - \xi^{(\ell)}|^2 | \bar{y}^{(\ell)} \} = W^2 \left(p^{(\ell)}(\cdot | \bar{y}^{(\ell)}), Q^{(\ell)}(\cdot | \bar{y}^{(\ell)}) \right) \quad \text{for any } \bar{y}^{(\ell)}.$$

By Assumptions 1 and 2, the Otto-Villani theorem can be applied to $Q^{(\ell)}(\cdot|\bar{y}^{(\ell)})$. Using also the logarithmic Sobolev inequality for $Q^{(\ell)}(\cdot|\bar{y}^{(\ell)})$, we get

$$\begin{aligned} \mathbb{E}_{\pi_\ell} \{ |\eta^{(\ell)} - \xi^{(\ell)}|^2 \mid \bar{y}^{(\ell)} \} &\leq \frac{2}{\rho_\ell} \cdot D(p^{(\ell)}(\cdot|\bar{y}^{(\ell)}) \parallel Q_i(\cdot|\bar{y}^{(\ell)})) \\ &\leq \frac{1}{\rho_\ell^2} \cdot I \left(p^{(\ell)}(\cdot|\bar{y}^{(\ell)}) \parallel Q^{(\ell)}(\cdot|\bar{y}^{(\ell)}) \right) = \frac{1}{\rho_\ell^2} \cdot \mathbb{E}_p \left\{ |\nabla^{(\ell)} \log h|^2 \mid \bar{y}^{(\ell)} \right\} \end{aligned} \quad (3.21)$$

for any $\bar{y}^{(\ell)}$. Substituting (3.21) into (3.20):

$$\begin{aligned} \mathbb{E}_{p,\Pi} \left| \left\{ \sum_{k,\ell \in [1,n], \ell \neq k} \rho_\ell \cdot [\nabla^{(k)} V(\bar{y}^{(\ell)}, \xi^{(\ell)}) - \nabla^{(k)} V(\bar{y}^{(\ell)}, \eta^{(\ell)})] \cdot \nabla^{(k)} \log h \right\} \right| \\ \leq \sum_{k=1}^n (\rho_k - \rho) \cdot \mathbb{E}_p |\nabla^{(k)} \log h|^2. \end{aligned} \quad (3.22)$$

Substituting (3.19) and (3.22) into (3.18):

$$\begin{aligned} \mathbb{E}_p \{ \nabla \chi \cdot \nabla \log h \} &\geq \sum_{k=1}^n \rho_k \cdot \mathbb{E}_p |\nabla^{(k)} \log h|^2 - \sum_{k=1}^n (\rho_k - \rho) \cdot \mathbb{E}_p |\nabla^{(k)} \log h|^2 \\ &= \rho \cdot \sum_{k=1}^n |\nabla^{(k)} \log h|^2 = \rho \cdot \mathbb{E}_p |\nabla \log h|^2. \end{aligned} \quad \square$$

4. Proof of the Auxiliary Theorem and the Approximation Lemma

In the proof of the Auxiliary Theorem we use the weighted Gibbs sampler Γ with weights ρ_k/R , $R = \sum_k \rho_k$, defined in Section 2:

$$\Gamma = \frac{1}{R} \cdot \sum_{k=1}^n \rho_k \cdot \Gamma_k, \quad \Gamma_k(z|y) = \delta(\bar{y}^{(k)}, \bar{z}^{(k)}) \cdot Q^{(k)}(z^{(k)}|\bar{y}^{(k)}).$$

Recall that

$$\sup_{x, \xi} \|A(x, \xi)\| \triangleq 1 - \delta < 1.$$

Proposition 4. *Under Assumptions 1-3, for fixed $z, u \in \mathbb{R}^N$ we have*

$$\begin{aligned} \sum_{k=1}^n \rho_k \cdot W^2\left(Q^{(k)}(\cdot|\bar{z}^{(k)}), Q^{(k)}(\cdot|\bar{u}^{(k)})\right) &\leq 2 \cdot \sum_{k=1}^n D\left(Q^{(k)}(\cdot|\bar{z}^{(k)}) \parallel Q^{(k)}(\cdot|\bar{u}^{(k)})\right) \\ &\leq (1 - \delta)^2 \cdot \sum_{k=1}^n \rho_k \cdot |z^{(k)} - u^{(k)}|^2. \end{aligned} \quad (4.1)$$

Proof.

The first inequality follows from the Otto-Villani theorem for $Q^{(k)}(\cdot|\bar{u}^{(k)})$. Then we use the logarithmic Sobolev inequality to continue (4.1) as follows:

$$\begin{aligned} &\leq \sum_{k=1}^n \frac{1}{\rho_k} \cdot I\left(Q^{(k)}(\cdot|\bar{z}^{(k)}) \parallel Q^{(k)}(\cdot|\bar{u}^{(k)})\right) \\ &= \int_{\mathbb{R}^N} \sum_{k=1}^n \frac{1}{\rho_k} \cdot \left| \nabla^{(k)} V(\bar{z}^{(k)}, \eta^{(k)}) - \nabla^{(k)} V(\bar{u}^{(k)}, \eta^{(k)}) \right|^2 \prod_{i=1}^n Q^{(k)}(d\eta^{(k)}|\bar{z}^{(k)}). \end{aligned} \quad (4.2)$$

To estimate the sum under the integral in (4.2), fix η^N , and consider the function $F = (F_1, \dots, F_N) : \mathbb{R}^N \mapsto \mathbb{R}^N$ defined by

$$\begin{aligned} F^{(k)} : \mathbb{R}^N &\mapsto \mathbb{R}^{(k)}, \\ F^{(k)}(z) &= \frac{1}{\sqrt{\rho_k}} \cdot \nabla^{(k)} V\left(\frac{z^{(1)}}{\sqrt{\rho_1}}, \dots, \frac{z^{(k-1)}}{\sqrt{\rho_{k-1}}}, \frac{\eta^{(k)}}{\sqrt{\rho_k}}, \frac{z^{(k+1)}}{\sqrt{\rho_{k+1}}}, \dots, \frac{z^{(n)}}{\sqrt{\rho_n}}\right). \end{aligned}$$

With the notation

$$\zeta^{(k)} = z^{(k)} \cdot \sqrt{\rho_k}, \quad \theta^{(k)} = u^{(k)} \cdot \sqrt{\rho_k}, \quad k \in [1, n],$$

the sum under the integral in (4.2) is just the squared increment of F between points ζ and θ :

$$\sum_{k=1}^n \frac{1}{\rho_k} \cdot \left| \nabla^{(k)} V(\bar{z}^{(k)}, \eta^{(k)}) - \nabla^{(k)} V(\bar{u}^{(k)}, \eta^{(k)}) \right|^2 = \sum_{k=1}^n \left| F^{(k)}(\zeta) - F^{(k)}(\theta) \right|^2. \quad (4.3)$$

The Jacobian of F is

$$\left(\frac{1}{\sqrt{\rho_k} \sqrt{\rho_\ell}} \cdot V_{i,j}(\bar{z}^{(k)}, \eta^{(k)}) \right)_{i \in I_k, j \in I_\ell, k \neq \ell}.$$

(It has zeros for i and j belonging to the same I_k .) Thus, by (2.4),

$$\sum_{k=1}^n |F_k(\zeta) - F_k(\theta)|^2 \leq (1 - \delta)^2 \cdot \sum_{k=1}^n \rho_k \cdot |z^{(k)} - u^{(k)}|^2. \quad (4.4)$$

Substituting (4.3) and (4.4) into (4.2) we get the desired result (4.1). \square

We use Proposition 4 to show that the Gibbs sampler Γ is a contraction with respect to a weighted Wasserstein distance.

Definition. Let r and s probability measures r and s on \mathbb{R}^N . We define the weighted quadratic Wasserstein distance of r and s (with wights ρ_k) by

$$W_{\{\rho_k\}}^2(r, s) = \inf_{\pi} \sum_{k=1}^n \rho_k \cdot \mathbb{E}_{\pi} |Z^{(k)} - U^{(k)}|^2,$$

where Z and U are random sequences s with laws r resp. s , and infimum is taken over all distributions $\pi = \mathcal{L}(Z, U)$ with marginals r and s .

Proposition 5.

If Assumptions 1-3 hold for q then

$$W_{\{\rho_k\}}(r\Gamma, s\Gamma) \leq \left(1 - \frac{\rho_{\min} \cdot \delta}{R}\right) \cdot W_{\{\rho_k\}}(r, s). \quad (4.5)$$

Proof.

Let $Z = (Z_1, Z_2, \dots, Z_N)$ and $U = (U_1, U_2, \dots, U_N)$ be random sequences in \mathbb{R}^N , with $\mathcal{L}(Z) = r$, $\mathcal{L}(U) = s$, and let $\pi = \mathcal{L}(Z, U)$ be that joining of r and s that achieves $W_{\{\rho_k\}}^2(r, s)$. Select a random index $\kappa \in [1, n]$ according to the distribution (ρ_k/R) , and define

$$\mathcal{L}(Z'|Z, U) = \Gamma_{\kappa}(\cdot|Z), \quad \mathcal{L}(U'|Z, U) = \Gamma_{\kappa}(\cdot|U).$$

Then $\mathcal{L}(Z') = r\Gamma$, and $\mathcal{L}(U') = s\Gamma$. Further, define $\mathcal{L}(Z', U')$ as that coupling of $r\Gamma$ and $s\Gamma$ that achieves $W^2(Q^{(\kappa)}(\cdot|\bar{Z}^{(\kappa)}), Q^{(\kappa)}(\cdot|\bar{U}^{(\kappa)}))$ for each value of the condition. Thereby we have defined $\mathcal{L}(Z', U'|Z, U)$, and by Proposition 4 we have

$$\begin{aligned} W_{\{\rho_k\}}^2(r\Gamma, s\Gamma) &\leq \sum_{k=1}^n \rho_k \cdot \left(1 - \frac{\rho_k}{R} + \frac{\rho_k}{R} \cdot (1 - \delta)^2\right) \cdot \mathbb{E}|Z^{(k)} - U^{(k)}|^2 \\ &= \left(1 - \frac{2\rho_{\min} \cdot \delta \cdot (1 - \delta/2)}{R}\right) \cdot \sum_{k=1}^n \rho_k \cdot \mathbb{E}|Z^{(k)} - U^{(k)}|^2 \\ &\leq \left(1 - \frac{\rho_{\min} \cdot \delta}{R}\right) \cdot W_{\{\rho_k\}}^2(r, s). \end{aligned} \quad \square$$

In the sequel we shall use the

Notation.

$$I(p^{(k)}(\cdot|\bar{Y}^{(k)}) \parallel Q^{(k)}(\cdot|\bar{Y}^{(k)})) \triangleq \mathbb{E}I(p^{(k)}(\cdot|\bar{Y}^{(k)}) \parallel Q_i(\cdot|\bar{Y}^{(k)}))$$

(omitting the symbol of expectation).

Proposition 6.

Under Assumptions 1-3 we have

$$\begin{aligned} W_{\{\rho_k\}}(p, q) &\leq \frac{2R}{\rho_{\min} \cdot \delta} \cdot \sqrt{\sum_{k=1}^n \rho_k \cdot \mathbb{E}W^2\left(p^{(k)}(\cdot|\bar{Y}^{(k)}), Q^{(k)}(\cdot|\bar{Y}^{(k)})\right)} \\ &\leq \frac{2R}{\rho_{\min} \cdot \delta} \cdot \sqrt{\sum_{k=1}^n \frac{1}{\rho_k} \cdot I\left(p^{(k)}(\cdot|\bar{Y}^{(k)}) \parallel Q^{(k)}(\cdot|\bar{Y}^{(k)})\right)}. \end{aligned}$$

Proof.

The first inequality follows from the triangle inequality for $W_{\{\rho_k\}}(p, q)$ and Proposition 5, and the second one follows from the Otto-Villani theorem and the logarithmic Sobolev inequality for $Q^{(k)}(\cdot|\bar{y}^{(k)})$. \square

Proposition 7.

There exists a $C = C(R, \rho_{\min}, \delta) > 0$ ($R = \sum_k \rho_k$ and $\rho_{\min} = \min_k \rho_k$) such that

$$\sum_{k=1}^n D(Y^{(k)} \parallel X^{(k)}) \leq \frac{1}{2C} \cdot I(p||q). \quad (4.6)$$

Proof.

Let $\pi = \mathcal{L}(Y, X)$ denote that joining of $p = \mathcal{L}(Y)$ and $q = \mathcal{L}(X)$ that achieves $W_{\{\rho_k\}}(p, q)$.

The convexity of the entropy functional implies the inequality

$$\sum_{k=1}^n D(Y^{(k)} \parallel X^{(k)}) \leq \sum_{k=1}^n \mathbb{E}_\pi D\left(Y^{(k)}|\bar{Y}^{(k)} \parallel Q^{(k)}(\cdot|\bar{X}^{(k)})\right). \quad (4.7)$$

The right-hand-side of (4.7) can be written as a sum of three terms:

$$\begin{aligned} &\sum_{k=1}^n \mathbb{E}_\pi D\left(Y^{(k)}|\bar{Y}^{(k)} \parallel Q^{(k)}(\cdot|\bar{X}^{(k)})\right) \\ &= \sum_{k=1}^n D\left(Y^{(k)}|\bar{Y}^{(k)} \parallel Q^{(k)}(\cdot|\bar{Y}^{(k)})\right) + \sum_{k=1}^n \mathbb{E}_\pi D\left(Q^{(k)}(\cdot|\bar{Y}^{(k)}) \parallel Q^{(k)}(\cdot|\bar{X}^{(k)})\right) \\ &+ \int_{\mathbb{R}^N \times \mathbb{R}^N} \sum_{k=1}^n \left[p^{(k)}(y^{(k)}|\bar{y}^{(k)}) - Q^{(k)}(y^{(k)}|\bar{y}^{(k)}) \right] \cdot \log \frac{Q^{(k)}(y^{(k)}|\bar{y}^{(k)})}{Q^{(k)}(y^{(k)}|\bar{x}^{(k)})} dy dx \\ &\triangleq S_1 + S_2 + S_3. \end{aligned} \quad (4.8)$$

By the logarithmic Sobolev inequality for $Q^{(k)}(\cdot|\bar{y}^{(k)})$ we have

$$S_1 \leq \frac{1}{2} \cdot \sum_{k=1}^n \frac{1}{\rho_k} I \left(Y^{(k)} | \bar{Y}^{(k)} \parallel Q^{(k)}(\cdot | \bar{Y}^{(k)}) \right). \quad (4.9)$$

Further, by Propositions 4,

$$S_2 \leq \frac{(1-\delta)^2}{2} \cdot \sum_{k=1}^n \rho_k \cdot \mathbb{E}_\pi |Y^{(k)} - X^{(k)}|^2 = \frac{(1-\delta)^2}{2} \cdot W_{\{\rho_k\}}(p, q). \quad (4.10)$$

S_3 can be written as

$$S_3 = \mathbb{E}_\mu \sum_{k=1}^n \left[V(Y) - V(\bar{X}^{(k)}, Y^{(k)}) - V(\bar{Y}^{(k)}, \xi^{(k)}) + V(\bar{X}^{(k)}, \xi^{(k)}) \right], \quad (4.11)$$

where $\mu = \mathcal{L}(Y, X, \xi)$ is defined by $\mathcal{L}(Y, X) = \pi$, $\mathcal{L}(\xi|Y, X) = \prod_{k=1}^n \mathcal{L}(\xi^{(k)}|Y)$, and $\mathcal{L}(Y^{(k)}, \xi^{(k)}|\bar{Y}^{(k)})$ is an arbitrary joining of $p^{(k)}(\cdot|\bar{Y}^{(k)})$ and $Q^{(k)}(\cdot|\bar{Y}^{(k)})$.

We claim that for any quadruple of sequences $(y^N, \eta^N, x^N, \xi^N)$ the following inequality holds:

$$\begin{aligned} & \sum_{k=1}^n \left[V(\bar{y}^{(k)}, \eta^{(k)}) - V(\bar{x}^{(k)}, \eta^{(k)}) - V(\bar{y}^{(k)}, \xi^{(k)}) + V(\bar{x}^{(k)}, \xi^{(k)}) \right] \\ & \leq (1-\delta) \cdot \sqrt{\sum_{k=1}^n \rho_k |y^{(k)} - x^{(k)}|^2} \cdot \sqrt{\sum_{k=1}^n \rho_k |\eta^{(k)} - \xi^{(k)}|^2}. \end{aligned} \quad (4.12)$$

Indeed, introducing the function

$$F : \mathbb{R}^N \times \mathbb{R}^N \mapsto \mathbb{R}, \quad F(y, \eta) = \sum_{k=1}^n V(\bar{y}^{(k)}, \eta^{(k)}),$$

the left-hand-side of (4.12) can be rewritten as follows:

$$\begin{aligned} & \sum_{k=1}^n \left[V(\bar{y}^{(k)}, \eta^{(k)}) - V(\bar{x}^{(k)}, \eta^{(k)}) - V(\bar{y}^{(k)}, \xi^{(k)}) + V(\bar{x}^{(k)}, \xi^{(k)}) \right] \\ & = F(y, \eta) - F(x, \eta) - F(y, \xi) + F(x, \xi). \end{aligned} \quad (4.13)$$

To estimate the right-hand-side of (4.13) (with $y, x, \eta, \xi \in \mathbb{R}^N$ fixed), define

$$\begin{aligned} G & : [0, 1] \times [0, 1] \mapsto \mathbb{R}, \\ G(s, t) & = F(x + s(y - x), \xi + t(\eta - \xi)) \\ & = \sum_{k=1}^n V(\bar{x}^{(k)} + s(\bar{y}^{(k)} - \bar{x}^{(k)}), \xi^{(k)} + t(\eta^{(k)} - \xi^{(k)})). \end{aligned}$$

Then we have

$$\begin{aligned} & F(y, \eta) - F(x, \eta) - F(y, \xi) + F(x, \xi) \\ &= G(1, 1) - G(1, 0) - G(0, 1) + G(0, 0). \end{aligned} \quad (4.14)$$

We have by (2.4)

$$\begin{aligned} & \left| \frac{\partial^2}{\partial_s \partial_t} G(s, t) \right| = \\ & \left| \sum_{k, \ell \in [1, n], k \neq \ell} \sum_{i \in I_k, j \in I_\ell} (y_i - x_i) \cdot V_{i,j}(\bar{x}^{(k)} + s(\bar{y}^{(k)} - \bar{x}^{(k)}), \xi^{(k)} + t(\eta^{(k)} - \xi^{(k)})) \cdot (\eta_j - \xi_j) \right| \\ & \leq (1 - \delta) \cdot \sqrt{\sum_{k=1}^n \rho_k \cdot |y^{(k)} - x^{(k)}|^2} \cdot \sqrt{\sum_{\ell=1}^n \rho_\ell \cdot |\eta^{(\ell)} - \xi^{(\ell)}|^2}. \end{aligned} \quad (4.15)$$

Putting together (4.13), (4.14) and (4.15) yields (4.12).

Applying (4.12) for $\eta = y$:

$$\begin{aligned} & \sum_{k=1}^n \left[V(y) - V(\bar{x}^{(k)}, y^{(k)}) - V(\bar{y}^{(k)}, \xi^{(k)}) + V(\bar{x}^{(k)}, \xi^{(k)}) \right] \\ & \leq (1 - \delta) \cdot \sqrt{\sum_{k=1}^n \rho_k |y^{(k)} - x^{(k)}|^2} \cdot \sqrt{\sum_{\ell=1}^n \rho_\ell |y^{(\ell)} - \xi^{(\ell)}|^2}. \end{aligned} \quad (4.16)$$

Substituting (4.16) into (4.11), and using Jensen's inequality, we get

$$\begin{aligned} S_3 & \leq \sqrt{\sum_{k=1}^n \rho_k \cdot \mathbb{E} |Y^{(k)} - X^{(k)}|^2} \cdot \sqrt{\sum_{k=1}^n \rho_k \cdot \mathbb{E} |Y^{(k)} - \xi^{(k)}|^2} \\ & = W_{\{\rho_k\}}(p, q) \cdot \sqrt{\sum_{k=1}^n \rho_k \cdot \mathbb{E} |Y^{(k)} - \xi^{(k)}|^2} \end{aligned} \quad (4.17)$$

To estimate the second factor, we select for $\mathcal{L}(Y^{(k)}, \xi^{(k)} | \bar{Y}^{(k)})$ that joining of the marginals that achieves $W^2(p^{(k)}(\cdot | \bar{Y}^{(k)}), Q^{(k)}(\cdot | \bar{Y}^{(k)}))$ for every value of the conditions. Then the Otto-Villani theorem and the logarithmic Sobolev inequality for $Q^{(k)}(\cdot | \bar{y}^{(k)})$ imply the following bound for S_3 :

$$S_3 \leq W_{\{\rho_k\}}(p, q) \cdot \sqrt{\sum_{k=1}^n \frac{1}{\rho_k} \cdot I \left(p^{(k)}(\cdot | \bar{Y}^{(k)}) || Q^{(k)}(\cdot | \bar{Y}^{(k)}) \right)}. \quad (4.18)$$

Putting together (4.9), (4.10) and (4.18):

$$\begin{aligned} & S_1 + S_2 + S_3 \\ & \leq \frac{1}{2} \cdot \left[W_{\{\rho_k\}}(p, q) + \sqrt{\sum_{k=1}^n \frac{1}{\rho_k} \cdot I\left(p^{(k)}(\cdot|\bar{Y}^{(k)}) || Q^{(k)}(\cdot|\bar{Y}^{(k)})\right)} \right]^2. \end{aligned} \quad (4.19)$$

(4.7), together with (4.8) and (4.19), completes the proof of Proposition 7. \square

Proof of the Auxiliary Theorem.

The proof goes by induction on n . It is clear that for any k and $y^{(k)} \in \mathbb{R}^{(k)}$, Assumptions 1-3 formulated before Theorem 1 do hold for $n = 1$, $N = |I_k|$ and the distribution $Q^{(k)}(\cdot|y^{(k)})$. Assume that we have proved the Auxiliary Theorem for $n - 1$ in place of n .

By a well known identity for relative entropy, we have

$$\begin{aligned} D(p||q) &= D(Y||X) \\ &= \frac{1}{n} \sum_{k=1}^n D(Y^{(k)} || X^{(k)}) + \frac{1}{n} \sum_{k=1}^n D\left(\bar{Y}^{(k)}|Y^{(k)} || \bar{q}^{(k)}(\cdot|Y^{(k)})\right). \end{aligned} \quad (4.20)$$

Assume the Auxiliary Theorem for $n - 1$. By the induction hypothesis,

$$D\left(\bar{Y}^{(k)}|Y^{(k)} || \bar{q}^{(k)}(\cdot|Y^{(k)})\right) \leq \frac{1}{2C} \cdot \sum_{\ell \neq k} I\left(Y^{(\ell)}|\bar{Y}^{(\ell)} || Q^{(\ell)}(\cdot|Y^{(\ell)})\right) \quad \text{for all } k.$$

Thus

$$\begin{aligned} & \frac{1}{n} \sum_{k=1}^n D(\bar{Y}^{(k)}|Y^{(k)} || \bar{q}^{(k)}(\cdot|Y^{(k)})) \\ & \leq (1 - 1/n) \cdot \frac{1}{2C} \sum_{\ell=1}^n I\left(Y^{(\ell)}|\bar{Y}^{(\ell)} || Q^{(\ell)}(\cdot|Y^{(\ell)})\right) \\ & = (1 - 1/n) \cdot \frac{1}{2C} \cdot I(p||q). \end{aligned} \quad (4.21)$$

Substituting (4.6) (Proposition 7) and (4.21) into (4.20) completes the proof of the Auxiliary Theorem. \square

Proof of the Approximation Lemma.

First we keep q fixed, and construct a density $g \in \mathcal{C}^\infty$ with compact support, and such that, with the notations

$$r = g \cdot q, \quad \bar{r}^{(k)} = \int_{\mathbb{R}^{(k)}} r(\bar{y}^{(k)}, \xi^{(k)}) d\xi^{(k)},$$

we have

$$D(r||q) \text{ is arbitrarily close to } D(p||q), \quad (4.22)$$

and

$$D(\bar{r}^{(k)}||\bar{q}^{(k)}) \text{ is arbitrarily close to } D(\bar{p}^{(k)}||\bar{q}^{(k)}) \text{ for all } k \in [1, n]. \quad (4.23)$$

Denote by B_m the closed ball in \mathbb{R}^N around the origin and with radius m . Let $\phi_m : \mathbb{R}^N \mapsto [0, 1]$ be a C^∞ function satisfying

$$\phi_m(x) = 1 \text{ for } x \in B_m, \quad \phi_m(x) = 0 \text{ for } x \notin B_{m+1}.$$

Set

$$g_m(x) = \frac{1}{\alpha_m} \cdot h(x) \cdot \phi_m(x), \quad \text{and} \quad r_m(x) = g_m(x) \cdot q(x),$$

where $\alpha_m = \int_{\mathbb{R}^N} h(x) \cdot \phi_m(x) q(dx)$.

We have

$$D(r_m||q) = \frac{1}{\alpha_m} \int_{\mathbb{R}^N} \left(h(x) \phi_m(x) \right) \cdot \log \left(h(x) \phi_m(x) \right) q(dx) - \log \alpha_m,$$

and $\lim_{m \rightarrow \infty} \alpha_m = 1$. Since $h(x) \cdot \phi_m(x) \rightarrow h(x)$ everywhere, with $|(h(x) \phi_m(x)) \cdot \log(h(x) \phi_m(x))|_+$ increasing, and using also the inequality $u \log u \geq -1/e$, it follows that

$$\lim_{m \rightarrow \infty} \int_{\mathbb{R}^N} \left(h(x) \phi_m(x) \right) \cdot \log \left(h(x) \phi_m(x) \right) q(dx) = D(p||q).$$

Putting $g = g_m$ and $r = r_m$, for large enough m we achieve (4.22). It can be proved similarly that (4.23) can be achieved as well.

Again, it is easily seen that

$$\lim_{\varepsilon \rightarrow 0} D\left((1 - \varepsilon) \cdot r + \varepsilon \cdot q \parallel q \right) = D(r||q),$$

and

$$\begin{aligned} & \lim_{\varepsilon \rightarrow 0} D\left((1 - \varepsilon) \cdot r^{(k)}(\cdot|\bar{Y}^{(k)}) + \varepsilon \cdot q^{(k)}(\cdot|\bar{Y}^{(k)}) \parallel q^{(k)}(\cdot|\bar{Y}^{(k)}) \right) \\ &= D(r^{(k)}(\cdot|\bar{Y}^{(k)}) \parallel q^{(k)}(\cdot|\bar{Y}^{(k)})). \end{aligned}$$

Thus, for q fixed, h can be replaced by $f = (1 - \varepsilon)g + \varepsilon$.

Now we can assume that h is of the form claimed in the Approximation lemma. We keep the notation $p = h \cdot q$ with the newly defined h , and keep h fixed.

Now we approximate $q(x)$ by an increasing sequence $\tilde{q}_m(x) \in \mathcal{C}^\infty$, and set

$$q_m = \frac{\tilde{q}_m}{\int \tilde{q}_m(x) dx} \quad \text{and} \quad \bar{q}_m^{(k)}(\bar{x}^{(k)}) = \int_{\mathbb{R}^{(k)}} q_m(x) dx^{(k)}.$$

Then define $p_m(x) = h(x) \cdot q_m(x)$. Since h is smooth and bounded from below and above, it is easily seen that

$$D(p_m || q_m) = \int_{\mathbb{R}^N} h(x) \log h(x) q_m(dx) \rightarrow D(p || q),$$

and

$$D(\bar{p}_m^{(k)} || \bar{q}_m^{(k)}) = \int_{\mathbb{R}^{(k)}} \bar{h}^{(k)}(\bar{x}^{(k)}) \cdot \log \bar{h}^{(k)}(\bar{x}^{(k)}) \bar{q}_m^{(k)}(d\bar{x}^{(k)}) \rightarrow D(\bar{p}^{(k)} || \bar{q}^{(k)}).$$

This completes the proof of the Approximation Lemma. □

Acknowledgment

The author is grateful for the patient support and help by P. E. Frenkel.

REFERENCES

- [1] L. Gross, *Logarithmic Sobolev inequalities*, Amer. J. Math. **97** (1975), 1061-1083.
- [2] E. Nelson, *The free Markov field*, J. Funct. Anal. **12** (1973), 211-277.
- [3] F. Otto, C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal. **173** (2000), 361-400.
- [4] Th. Bodineau, B. Helffer, *On Log-Sobolev inequalities for unbounded spin systems*, Funct. Anal. **166**, (1999), 168-178.
- [5] B. Helffer, *Remarks on decay of correlation and Witten-Laplacians III. Application to logarithmic Sobolev inequalities*, Ann. Inst. H. Poincaré, **35**, (1999), 483-508.
- [6] D. Bakry, M. Emery, *Diffusions hypercontractives*, Séminaire de Probabilités, XIX. Lecture Notes in Math. 1123, Springer, 1985, pp. 177-206.
- [7] R. Holley, D. Stroock, *Logarithmic Sobolev inequalities and stochastic Ising models*, J. Statist. Phys. **16** (1987), 1159-1191.
- [8] B. Zegarlinski, *Dobrushin's uniqueness theorem and logarithmic Sobolev inequalities*, J. Funct. Anal. **105** (1992), 77-111.
- [9] G. Royer, *Une Initiation aux Inégalités de Sobolev Logarithmiques*, Soc. Math. de France, 1999.
- [10] F. Otto, M. Reznikoff, *A new criterion for the logarithmic Sobolev inequality and two applications*, J. Funct. Anal. **243**, (2011), 121-157.
- [11] M. Talagrand, *Transportation cost for Gaussian and other product measures*, Geometric and Functional Analysis, **6** (1996), 587-600.
- [12] S. Bobkov, I. Gentil, M. Ledoux, *Hypercontractivity of Hamilton-Jacobi equations*, J. Math. Pures Appl. **80** (2001), 669-696.
- [13] K. Marton, *Measure concentration for Euclidean distance in the case of dependent random variables*, Ann. Prob. **32** (2004), 2526-2544.

- [14] K. Marton, *Correction to the paper "Measure concentration for Euclidean distance in the case of dependent random variables"*, Annals of Probability . **38** (2010), 438-442.
- [15] U. Grenander, G. Szegő, *Toeplitz forms and their applications*, University of California Press, Berkley, 1958.
- [16] F. Riesz, B. Szőkefalvi-Nagy, *Functional analysis*, Ungar, New York, 1955.

H-1364 POB 127, BUDAPEST, TEL. 36 (1) 4838300, HUNGARY
E-mail address: `marton@renyi.hu`